Aula

5

COLETA, TRATAMENTO E ANÁLISE DA INFORMAÇÃO GEOGRÁFICA

META

Conhecer as bases estatísticas para a tradução geográfica através da coleta, tratamento e análise da informação geográfica

OBJETIVOS

Ao final desta aula o aluno deverá:
classificar e localizar as fontes de coleta de dados;
identificar as séries estatísticas;
tratar dados geográficos;
sistematizar dados quantitativos.

PRÉ-REQUISITO

Revisar as operações matemáticas básicas: soma, subtração, multiplicação e divisão, que servirão de base para o tratamento e a sistematização dos dados geográficos.

INTRODUÇÃO

Microsoft

A citação de marca comercial não implica na recomendação do referido programa, tendo sido feita apenas devido à facilidade de encontrar o Excel em qualquer computador.

Na aula anterior, aprendemos alguns processos fundamentais para a elaboração de mapas temáticos, como identificar, coletar e utilizar fonte de dados alfanumérica e o mapa-base, além de utilizar adequadamente os elementos cartográficos indispensáveis à elaboração desses mapas.

Nesta quinta aula iremos nos aprofundar um pouco nas bases que outras ciências nos dão para facilitar a coleta, tratamento e análise dos dados geográficos. Estudaremos um pouco de Estatística, somente no que se refere às regras de arredondamento e sistematização de dados quantitativos, através das medidas de tendência central. Envolverá um pouco de Matemática, mas cujos cálculos poderão ser facilitados utilizando-se uma calculadora ou uma planilha de cálculo, como a Excel da **Microsoft**.

BASES ESTATÍSTICAS PARA A TRADUÇÃO GRÁFICA

Os dados têm sua classificação baseada na fonte de coleta, que pode ser primária ou secundária.

- Fonte primária: dados, fatos, testemunhos originais relativos a uma situação. Em Geografia existe uma diversidade de fontes primárias, como registros de *trabalho de campo, entrevistas, levantamento cartográfico, livros, artigos, anuários estatísticos, entre outros*.

Cada trabalho tem sua especificidade. Os dados podem ter como fonte primária, tanto o próprio autor do trabalho no caso de uma pesquisa de campo, como uma instituição oficial como o IBGE, quando o tema estudado é a quantidade de população, por exemplo. Já os dados ambientais devem ser preferencialmente coletados em órgãos públicos como IBAMA ou as instituições ambientais estaduais e municipais, até mesmo as universidades, a depender do caso.

Quando nos referimos a fontes primárias que se encontram na forma de livros ou artigos, é mais profícuo ler, registrar e analisar livros ou artigos de autores considerados clássicos ou peritos no assunto.

- Fonte secundária: são **os** dados originais inseridos e/ou tratados em outras publicações. No caso de um livro, poderia ser um tema discutido pelo autor, mas referindo-se a uma publicação de outro autor. Uma tradução também é considerada uma fonte secundária, pois por mais respeitado que seja o tradutor e por mais fiel que seja a tradução, é a leitura de alguém sobre a obra original.

Para os trabalhos em Geografia, deve-se preferir trabalhar com as fontes primárias, utilizando-se das secundárias apenas quando não exista a possibilidade do uso das primeiras ou então que seja inviável por questões financeiras ou de tempo.

Isso não quer dizer que, se encontrarmos apenas o livro traduzido ou uma bibliografia inicial que seja mais baseada em fontes secundárias, devamos desistir do projeto. Enquanto não conseguimos as fontes primárias devemos trabalhar com as secundárias, mas sem perder de vista essa necessidade.

As implicações resultantes devem ser analisadas cuidadosamente pelo geógrafo de forma a não resultar em trabalhos de qualidade ruim ou duvidosa.

A localização dos dados poderá ser feita principalmente nas bibliotecas especializadas na área de Geografia ou afins. Em Aracaju temos a biblioteca do IBGE, da SEPLAN Estadual, e outras, além da própria UFS, mas a depender do tema a ser estudado, você poderá encontrar dados também nos municípios.

Importante atualmente é a pesquisa pela Internet. No entanto, devemos ter um cuidado especial ao utilizar fontes nesta mídia. Recomendamos apenas a utilização de *sites* oficiais, ou seja, de órgãos públicos, entidades de ensino ou pesquisa que sejam responsáveis oficiais pelo dado produzido. Também existem muitas revistas especializadas em Geografia e em outras áreas que podem ser consultadas *on line*, como Terra Livre e Geousp, ambas avaliadas pelo Qualis como periódicos nacionais conceito A, oferecendo na rede os artigos publicados. Nos exemplos citados, a Internet pode ser uma boa fonte de consulta.

ORDENAMENTO DOS DADOS (SÉRIES ESTATÍSTICAS)

Realizada a pesquisa nas fontes, é preciso agora ordenar os dados, de acordo com a necessidade do desenvolvimento da pesquisa ou do tema a ser estudado. Esses dados são ordenados segundo as séries estatísticas.

Uma Série Estatística corresponde a um quadro ou tabela em que se resume um conjunto de observações segundo o tempo, o espaço ou a categoria do fato, resultando, respectivamente nas Séries Históricas, Geográficas, Categóricas ou Múltiplas. Uma série estatística é constituída de linhas, colunas e células (encontro de uma linha com uma coluna).

- Série Histórica: mostra a variação de um fato através do tempo (semanas, meses, anos, etc.). Também é chamada de Cronológica ou Temporal. Nesta série, o local e a categoria são fixos, enquanto que o tempo é que varia (Tabela 1).

Ano	População (1.000 habitantes)
1950	51944
1960	70191
1970	93139

- Série Geográfica: mostra a variação ou ocorrência de um fato através do espaço. Neste tipo, o tempo e a categoria são fixos, variando o espaço, ou seja, o local de ocorrência (Tabela 2).

Tabela 5.2 – Estado de Sergipe – Percentual da área dos Territórios Sergipanos - 2008.

Nome do Território	%
Alto Sertão Sergipano	22,37
Baixo São Francisco Sergipano	8,88
Médio Sertão Sergipano	7,22
Leste Sergipano	6,93
Agreste Central Sergipano	14,25
Sul Sergipano	14,29
Centro-Sul Sergipano	16,07
Grande Aracaju	9,98

Tempo fixo (2008); espaço variável (territórios sergipanos); categoria fixa (% da área).

Fonte: Mapa dos Territórios Sergipanos, 2008.

- Série Categórica: mostra o fato por categorias ou espécie. Neste caso, local e tempo são fixos, variando a categoria (Tabela 5.3).

Tabela 5.3 – Estado de Sergipe – Efetivo da Pecuária – 2004

Rebanho	Número de cabeças
Bovino	929.794
Ovino	139.063
Caprino	15.130
Equino	68.640
Suíno	101.481
Ave	4.555.372

Tempo fixo (2004); espaço fixo (Estado de Sergipe); categoria variável (tipo de rebanho).
Fonte: IBGE (2004) e FRANÇA (2007)

- Série Múltipla: há momentos em que as tabelas combinam as séries vistas anteriormente, sendo denominadas então de Série de Múltipla Entrada ou simplesmente Série Múltipla (Tabela 5.4).

Tabela 5.4 – Grande Aracaju – Unidades Habitacionais construídas pela COHAB – 1968 a 2002

Município	Tipo de unidade habitacional	
	Casa	Apartamento
Aracaju	17.291	3.296
Barra dos Coqueiros	916	-
Nossa Senhora do Socorro	17.433	-
São Cristóvão	3.813	-

Tempo fixo (ano); espaço variável (município); categoria variável (tipo de unidade habitacional).
Fonte: CEHOP (2003) e FRANÇA (2007).

TRATAMENTO DOS DADOS

Uma vez coletados os dados e ordenados em séries estatísticas, eles estão prontos para serem trabalhados com a Estatística. Porém, geralmente requerem antes, um tratamento de forma a deixá-los compatíveis com as unidades de estudo. Ao ser feita uma simples média de crianças vacinadas contra a poliomielite em relação ao número de habitantes total do Estado de Sergipe, por exemplo, o resultado não deve conter casas decimais, pois a unidade mínima para habitantes é o inteiro. Para tanto, precisaremos realizar o arredondamento dos números para o número de casas decimais necessárias ou compatíveis com o fenômeno geográfico em estudo.

TÉCNICA DE ARREDONDAMENTO

Há casos em que o arredondamento de dados é plenamente justificável, visto não alterar a substância ou essência do trabalho, especialmente quando se sabe que os dados não são precisos e, muito principalmente, quando a facilidade do tratamento das informações é mais importante do que a precisão. De modo geral, também se justifica o arredondamento quando os números são resultantes de diversos cálculos que levam a certa generalização, como médias e porcentagens, por exemplo. Para fazermos o arredondamento, devemos adotar o seguinte critério (Tabela 5.5)
a) O algarismo a permanecer não sofrerá alteração alguma caso o que lhe seja imediatamente posterior seja menor que 5, ou seja, 1, 2, 3 ou 4. Exemplo: ao arredondarmos 12,4 para inteiros, o resultado será 12 já que o algarismo a ser abandonado é o 4.
b) O último dos algarismos a permanecer será aumentado de uma unidade caso o que lhe seja imediatamente posterior seja maior que 5, ou seja, 6, 7, 8 ou 9. Exemplo: ao arredondarmos 12,7 para inteiros, o resultado será 13 já que o algarismo a ser abandonado é o 7.
c) Se o algarismo a abandonar for exatamente igual a 5, porém seguido de outros números, isto o torna maior que 5; então o arredondamento far-se-á conforme a regra anterior. Exemplo: ao arredondarmos 12,514 para inteiros, o resultado será 13.
d) Se o algarismo a abandonar for exatamente igual a 5 sem quaisquer outros números a sua direita ou então seguido somente de zero(s), o algarismo a ser conservado será aumentado de uma unidade se for ímpar, permanecendo sem alterações caso seja par. Exemplos: 11,5 $\approx$ 12; 12,5 $\approx$ 12; 13,5 14; 12,50 $\approx$ 13; 13,50 $\approx$ 14.

A prática do arredondamento faz com que, ao longo das operações, os aumentos e reduções devidos aos arredondamentos sejam compensados.

Tabela 5.5: Critérios de arredondamento (Resolução nº 866/66 – IBGE).

Condições (resto)	Procedimentos	Exemplos
< 5	O último algarismo a permanecer fica inalterado.	12,4 passa para 12
> 5	Aumenta-se de uma unidade o algarismo a permanecer.	12,7 passa para 13
= 5	(i) Se ao 5 seguir em qualquer casa um algarismo diferente de zero, aumenta-se uma unidade no algarismo a permanecer.	12,514 passa para 12
	(ii) Se o 5 for o último algarismo ou se ao 5 só seguirem zeros, o último algarismo a ser conservado só será aumentado de uma unidade se for ímpar.	11,5 passa para 12 12,5 passa para 12 13,5 passa para 14 12,50 passa para 13 13,50 passa para 14

Um cuidado especial deve ser tomado ao utilizar planilhas eletrônicas como, por exemplo, o Excel da Microsoft. Quando o algarismo imediatamente posterior ao algarismo que deverá permanecer for igual a 5, o software não utiliza a regra dos números antecedentes par/ímpar, arredondando sempre para mais. Isto acarreta maior desequilíbrio na média dos arredondamentos, ampliando as possibilidades da soma dos valores após o arredondamento não coincidir com o total pretendido. Exemplos: no Excel: 12,5 = 13 (Diferente da regra); 13,5 = 14 (Igual à regra).

Neste caso acarretará um desequilíbrio na compensação dos dados, recomendando-se verificar se os resultados trarão ou não prejuízos para a análise do tema em estudo.

DÍGITOS SIGNIFICATIVOS

A precisão de medidas de dados contínuos sempre pode ser aprimorada melhorando-se o instrumento de medida. Por isso, os estatísticos fazem distinção entre dígitos significativos, que representam uma informação precisa, e dígitos que servem apenas para localizar a vírgula. Nos resultados, devem ser apresentados apenas os dígitos significativos, para evitar a falsa impressão de exatidão (Vieira, 1999). O resultado de um cálculo estatístico não deve conter mais dígitos significativos que os dados de menor precisão.

Por exemplo, um sensor de temperatura (termômetro de mercúrio) tem uma precisão de 0,2°C. Isso significa que as leituras com precisão serão obtidas em intervalos de 0,2 em 0,2°C e, por extrapolação, poderia se chegar a leituras intermediárias de 0,1 (Tabela 6). Ao final, teremos uma tabela com intervalos de 0,2 ou 0,1ºC. A precisão centesimal não tem significado numérico algum para estas medidas, pois a precisão de medidas é de no máximo 0,1°C. Deve-

se reduzir a precisão para o número de casas decimais compatíveis com a precisão do sensor que gerou o valor, ou seja, 21,1ºC.

Tabela 5.6: Exemplo hipotético de valores de temperatura do ar.

Hora	Temperatura do ar (ºC)
06h00min	18,2
08h00min	19,5
10h00min	20,6
12h00min	22,9
14h00min	24,5
Média	**21,14 ≈ 21,1**

(Fonte: GALVANI, 2005).

Muitas vezes, após o arredondamento, a soma dos valores não coincide com o total pretendido. Quando isto ocorre, é recomendado adicionar ou subtrair, conforme o caso, o valor faltante ou excedente à maior parcela ou então àquela que se considerar que o erro relativo será menor. Além disso, podemos também distribuir o excesso ou falta pelas parcelas, usando o bom senso, de modo que não ocorra significativa alteração dos dados.

SISTEMATIZAÇÃO DE DADOS QUANTITATIVOS

(Texto extraído de GALVANI, 2005)

A graduação em Geografia exige, pela própria natureza do curso, um número significativo de trabalhos de campo. Essas *saídas* realizadas pelas diferentes áreas/disciplinas, cada qual com seu instrumental apropriado, produzem em cada trabalho um volume de informações. Quando retomamos para a sala de aula, a grande questão que se apresenta é *"o que fazer com os dados quali-quantitativos coletados no trabalho de campo?"* Tradicionalmente, os alunos de graduação em Geografia não são muito afeitos à área de exatas o que, em certas ocasiões, limita a análise e interpretação dos dados observados. Por vezes, os resultados finais obtidos ficam prejudicados por falta de uma análise mais numérica (estatística) dos dados observados. O que pretendemos aqui não é formar especialistas em Estatística, mas fornecer alguns princípios básicos de Estatística Descritiva que permitam uma melhor análise dos dados obtidos nos trabalhos de campo e, também, desmistificar o trauma que é imposto aos nossos alunos com relação às ciências exatas. Vale lembrar que essa sistematização de dados quantitativos, como chamaremos a partir deste momento, aplica-se a qualquer tipo de informação, seja ela produto de questionários ou de medições específicas em cada área/ disciplina do conhecimento (GALVANI, 2008).

MEDIDAS DE TENDÊNCIA CENTRAL

A análise de um conjunto de dados com uso de tendência central nos permite avaliar para onde caminha nosso dado. Uma espécie de *raios-X* inicial que pode ser determinado com a utilização dos indicadores descritos a seguir.

Média Aritmética ($\overline{X}$): a média aritmética é o procedimento mais simples e comum passível de ser aplicado a um conjunto de dados. Esta medida de tendência central expressa o somatório de todos os elementos da série dividido pelo número total de elementos. Numericamente, a média aritmética é expressa por:

$$\overline{X} = \frac{\sum_{1}^{n} X_i}{n}$$

X_i é cada elemento da série, e *i* varia de 1 a *n; n* é o número de elementos e o símbolo Ó significa somatório de todos os elementos da série. Resumindo, somam-se todos os elementos e divide-se pelo número total de elementos da série.

Moda (MO): a moda ou modo (MO) é o valor presente na série que ocorre com mais freqüência. Existem séries em que nenhum dado se repete, nesses casos não existe a moda da série. Isso geralmente ocorre em séries reduzidas (menos que cinquenta elementos amostrados). De forma análoga, podem ocorrer séries com duas (bi-modal) ou mais modas. Nesses casos prevalece o valor de maior freqüência de ocorrência ou, em caso de empate, a série pode apresentar mais de uma moda.

Mediana (ME): a mediana é aplicável em séries extensas de dados (geralmente mais de mil informações nas quais existem extremos que possam *contaminar* a média, ou seja, alguns dados que *fogem* da *tendência central,* podendo sub ou superestimar as análises. A mediana é determinada ordenando-se os dados de forma crescente ou decrescente e identificando-se a posição central da série. Em caso de séries com número impar de elementos, a mediana estará na posição central da série. Para séries com número par de elementos, a mediana será a média dos elementos que ocupam a posição central da série. O conceito de mediana gera algumas confusões: a mediana é simplesmente o valor que se situa na posição central do conjunto de dados ordenados. Assim, tem que haver uma relação de ordem nos valores.

Valor Máximo (V_{max}) e Valor Mínimo (V_{min}): o valor máximo da série é aquele de maior magnitude, ou seja, o maior valor encontrado na série. O valor mínimo, por sua vez, é o menor valor encontrado na série. Em princípio, parece ser uma informação sem importância. Contudo nos permite visualizar em que intervalo de medidas encontra-se distribuído determinado conjunto de dados. Serve para evidenciar o *tamanho* dos da-

dos que serão trabalhados. Em séries climatológicas de temperatura do ar, por exemplo, o Vmax equivale à temperatura máxima do ar e o Vmin equivale à altura mínima do ar.

Amplitude (Ä): a amplitude de um conjunto de dados expressa a diferença entre o Vmax e o Vmin. Essa medida de tendência central expressa a variação máxima dos valores constituintes do conjunto de dados. Dois ou mais conjuntos de dados poderão ter a mesma média, porém diferentes Vmax, Vmin e Ä , evidenciando-se tratar de séries distintas. A seguir será apresentado, para um conjunto simples de dados, exemplo de cálculo das medidas de tendência central: média, moda, mediana, valor máximo, valor mínimo e amplitude.

Esses procedimentos podem ser efetuados *facilmente* com o programa *Excel* da Microsoft através dos seguintes passos: dispor o conjunto de dados em duas colunas; entrar na barra de ferramentas no atalho *fx;* em seguida em *estatística* e selecionar a análise de tendência desejada; selecionar o intervalo de dados. O resultado é mostrado. Caso a barra de ferramentas não ofereça o atalho *fx,* clique em *inserir* e em seguida em *fx,* seguindo os mesmos procedimentos descritos acima.

Qual a diferença de interpretação entre a mediana e a média?

Embora a média seja um valor mais fácil de ser entendido, seu uso tem restrições na medida em que pode nos induzir a um erro de tendência, se a amostra analisada apresentar valores de amplitude elevados. Por exemplo, na distribuição dos dados da Tabela 5.7 a média da variável A é 161 e a mediana é 163. Caso a amostra tivesse apresentado valor máximo de 300 e não 184, a média saltaria para 178, ou seja, seria superior a todos os valores individuais, mas a mediana continuaria a ser 163. Se olharmos para os 7 valores individuais da nossa amostra, verificamos que o número 163 é o melhor representante da distribuição desse conjunto de dados. Assim, no caso das variáveis quantitativas, quando o valor da mediana é muito diferente da média, é aconselhável considerar sempre a mediana como valor de referência mais importante.

Tabela 5.7 – Valores arbitrários para duas variáveis A e B.

Dados brutos			Dados ordenados (crescente)	
A	B		A	B
121	171		121	152
171	152		157	168
158	170		158	169
173	168		163	170
184	169		171	171
163	171		173	171
157	190		184	190

Tabela 5.8 – Resultados da análise de tendência central para o conjunto de dados da tabela anterior.

Medida de tendência	Variável A	Variável B
Média	161	170
Moda	?	171
Mediana	163	170
Valor Máximo	181	190
Valor Mínimo	121	152
Amplitude	63	38

MEDIDAS DE DISPERSÃO, DISTRIBUIÇÃO DE FREQUÊNCIA E CORRELAÇÃO E REGRESSÃO LINEAR

Estas formas de sistematização de dados requerem um conhecimento mais aprofundado em Estatística e não serão exigidas nas atividades e nas avaliações. No entanto, você poderá consultar uma bibliografia específica ou então a referência bibliográfica indicada nesta aula (GALVANI, 2005). As medidas de dispersão são úteis quando diferentes conjuntos de dados apresentam a mesma média e mediana, porém variabilidades distintas. Este tipo de análise pode ser utilizado para comparar quantos conjuntos de dados forem necessários, pois os cálculos são efetuados individualmente para cada conjunto. Subdividem-se em: *desvio* em relação à média, variância, desvio padrão e coeficiente de variação.

A distribuição de frequência é composta pela frequência, frequência relativa, probabilidade e tempo de retorno.

A correlação e regressão linear são determinadas pelo coeficiente de correlação e pela regressão linear.

CONCLUSÃO

A coleta, tratamento e análise da informação geográfica são muito importantes para os estudos geográficos, mas requerem alguns cuidados e conhecimentos básicos de Estatística. Conhecer as bases estatísticas como séries, arredondamento de dados e medidas de tendência central são fundamentais para uma posterior tradução gráfica dos dados.

RESUMO

Nesta aula, aprendemos que a Estatística é uma ciência que fornece aporte à Geografia no tocante ao tratamento e análise da informação. Iniciamos classificando e localizando as principais fontes de coleta de dados, relacionando inclusive alguns órgãos públicos que contêm bibliotecas (inclusive as virtuais) com publicações ou dados utilizados na Geografia. Em seguida aprendemos como os dados são dispostos, através das séries estatísticas históricas, geográficas, categóricas ou múltiplos. Depois estudamos as formas de tratamento e sistematização dessas séries, utilizando as regras de arredondamento e as medidas de tendência central, conceitos básicos da Estatística.

ATIVIDADES

1. Forneça um exemplo real da cada uma das Séries Estatísticas. Não se esqueça de indicar o título e a fonte de consulta.
2. Para cada uma das séries da atividade anterior, calcule todas as Medidas de Tendência Central, observando sempre as regras de arredondamento de forma a deixar o resultado com valores compatíveis com o elemento geográfico existente na tabela.

COMENTÁRIO SOBRE AS ATIVIDADES

Na primeira atividade, você poderá consultar as bibliotecas indicadas no texto referente às Fontes de Dados. Cuidado com as consultas na Internet! Lembre-se de utilizar apenas fontes oficiais. Para a segunda, de preferência utilize uma calculadora científica ou então uma planilha de cálculo, como a Excel. Lembre-se que todos os resultados devem ser arredondados para o mesmo número de casas decimais da fonte de dados.

PRÓXIMA AULA

Na próxima aula, veremos como transformar dados quantitativos em gráficos, que são muito utilizados na Geografia e também na Cartografia quando da elaboração de cartodiagramas, que são mapas-base com a inserção de diagramas (gráficos), evidenciando um caráter quantitativo ou dinâmico ao mapa.

REFERÊNCIAS

ALVES, Glória da Anunciação; CUSTÓDIO, Vanderli. *P*esquisa Bibliográfica e Fonte de Dados. In: VENTURI, Luís Antonio Bittar (Org.). **Praticando Geografia:** técnicas de campo e laboratório. São Paulo: Oficina de Textos, 2005. p.203-210.

FRANÇA, Vera Lúcia; CRUZ, Maria Tereza Souza (Org.). **Atlas escolar de Sergipe:** espaço geo-histórico e cultural. João Pessoa: GRAFSET, 2007.

GALVANI, Emerson. Sistematização de sados quantitativos. In: VENTURI, Luís Antonio Bittar (Org). **Praticando Geografia:** técnicas de campo e laboratório. São Paulo: Oficina de Textos, 2005. p.175-186.

OLIVEIRA, Paulo José de. **Cartografia Temática.** Apostila. São Cristóvão: UFS, 2008.